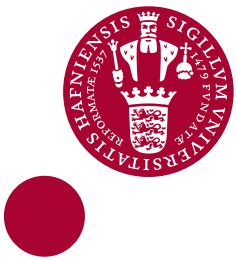


Single-Cell Gene Expression Analysis of a Human ESC Model of Pancreatic Endocrine Development Reveals Different Paths to -Cell Differentiation

Petersen, Maja Borup Kjær; Azad, Ajuna; Ingvorsen, Camilla; Hess, Katja; Hansson, Mattias; Grapin-Botton, Anne; Honoré, Christian

Published in:
Stem Cell Reports

DOI:
10.1016/j.stemcr.2017.08.009

Publication date:
2017

Document version
Publisher's PDF, also known as Version of record

Stem Cell Reports
Article

ISSCR

OPEN ACCESS

Single-Cell Gene Expression Analysis of a Human ESC Model of Pancreatic Endocrine Development Reveals Different Paths to β-Cell Differentiation

Maja Borup Kjær Petersen,[1,2] Ajuna Azad,[2] Camilla Ingvorsen,[3] Katja Hess,[2] Mattias Hansson,[4] Anne Grapin-Botton,[2,*] and Christian Honoré[1,*]

[1]Department of Stem Cell Biology, Novo Nordisk A/S, Novo Nordisk Park, 2760 Måløv, Denmark
[2]DanStem, University of Copenhagen, 3B Blegdamsvej, 2200 Copenhagen N, Denmark
[3]Histology and Imaging, Novo Nordisk A/S, Novo Nordisk Park, 2760 Måløv, Denmark
[4]Global Research External Affairs, Novo Nordisk A/S, Novo Nordisk Park, 2760 Måløv, Denmark
*Correspondence: anne.grapin-botton@sund.ku.dk (A.G.-B.), clfh@novonordisk.com (C.H.)
http://dx.doi.org/10.1016/j.stemcr.2017.08.009

SUMMARY

The production of insulin-producing β cells from human embryonic stem cells (hESCs) *in vitro* represents a promising strategy for a cell-based therapy for type 1 diabetes mellitus. To explore the cellular heterogeneity and temporal progression of endocrine progenitors and their progeny, we performed single-cell qPCR on more than 500 cells across several stages of *in vitro* differentiation of hESCs and compared them with human islets. We reveal distinct subpopulations along the endocrine differentiation path and an early lineage bifurcation toward either polyhormonal cells or β-like cells. We uncover several similarities and differences with mouse development and reveal that cells can take multiple paths to the same differentiation state, a principle that could be relevant to other systems. Notably, activation of the key β-cell transcription factor NKX6.1 can be initiated before or after endocrine commitment. The single-cell temporal resolution we provide can be used to improve the production of functional β cells.

INTRODUCTION

The generation of β cells from human pluripotent stem cells (hPSCs) receives much attention due to the potential for developing a cell-based therapy for diabetes mellitus and for use in drug discovery and disease modeling *in vitro*. Considerable progress has been made in the development of protocols for *in-vitro*-directed differentiation of hPSCs toward β cells. While current protocols produce β-like cells with some functionality (Pagliuca et al., 2014; Rezania et al., 2014; Russ et al., 2015), the generated cells are not fully mature and possess limited glucose responsiveness (Johnson, 2016).

The design of stepwise directed differentiation protocols to produce islet-like cells has relied heavily on knowledge from pancreas developmental biology. While it is well established that the five types of endocrine cells found in the islets of the pancreas (α, β, δ, ε, and PP cells) are derived from a progenitor population characterized by transient expression of *NEUROG3* (Gu et al., 2002), it remains unknown how the individual endocrine cell types are segregated from this population. However, several studies in mice and *in vitro* human embryonic stem cell (hESC) differentiation suggest that β cells differentiate from a subset of pancreatic progenitors expressing PDX1$^+$ and NKX6.1$^+$ that will turn on NEUROG3 (Kelly et al., 2011; Kroon et al., 2008; Nelson et al., 2007; Rezania et al., 2013; Schaffer et al., 2013).

In aiming to achieve the goal of producing stem cell-derived fully functional β cells that closely resemble human primary β cells, the need for a deeper phenotyping of both human β cells and stem cell-derived cultures has been emphasized (Johnson, 2016). Single-cell RNA sequencing (RNA-seq) has recently been applied to characterize single human islet cells (Baron et al., 2016; Lawlor et al., 2017; Li et al., 2015; Muraro et al., 2016; Segerstolpe et al., 2016; Wang et al., 2016; Xin et al., 2016), but single-cell gene expression profiling of hPSC-derived cultures differentiated toward the pancreatic lineage has, to the best of our knowledge, not been reported. Single-cell-based analysis offers the potential to reveal heterogeneity in differentiated hPSC cultures that can affect the propensity to differentiate into specific cell types.

To do so, we studied the formation of pancreatic endocrine cells *in vitro* using a model system based on differentiation of hESCs toward the pancreatic endocrine lineage. We analyzed more than 500 cells isolated from several stages of differentiation by single-cell qPCR and compared them with primary human islet cells. The low noise of single-cell PCR enabled us to establish a transcriptional map of the progressive stages of differentiation during endocrine development and uncovered prospective lineage trees for cells fated to become either polyhormonal or β-cell like. Integration of single-cell gene expression analysis with functional studies revealed multiple differentiation paths to β-like cells through the initiation of NKX6.1 expression either before or after endocrine commitment.

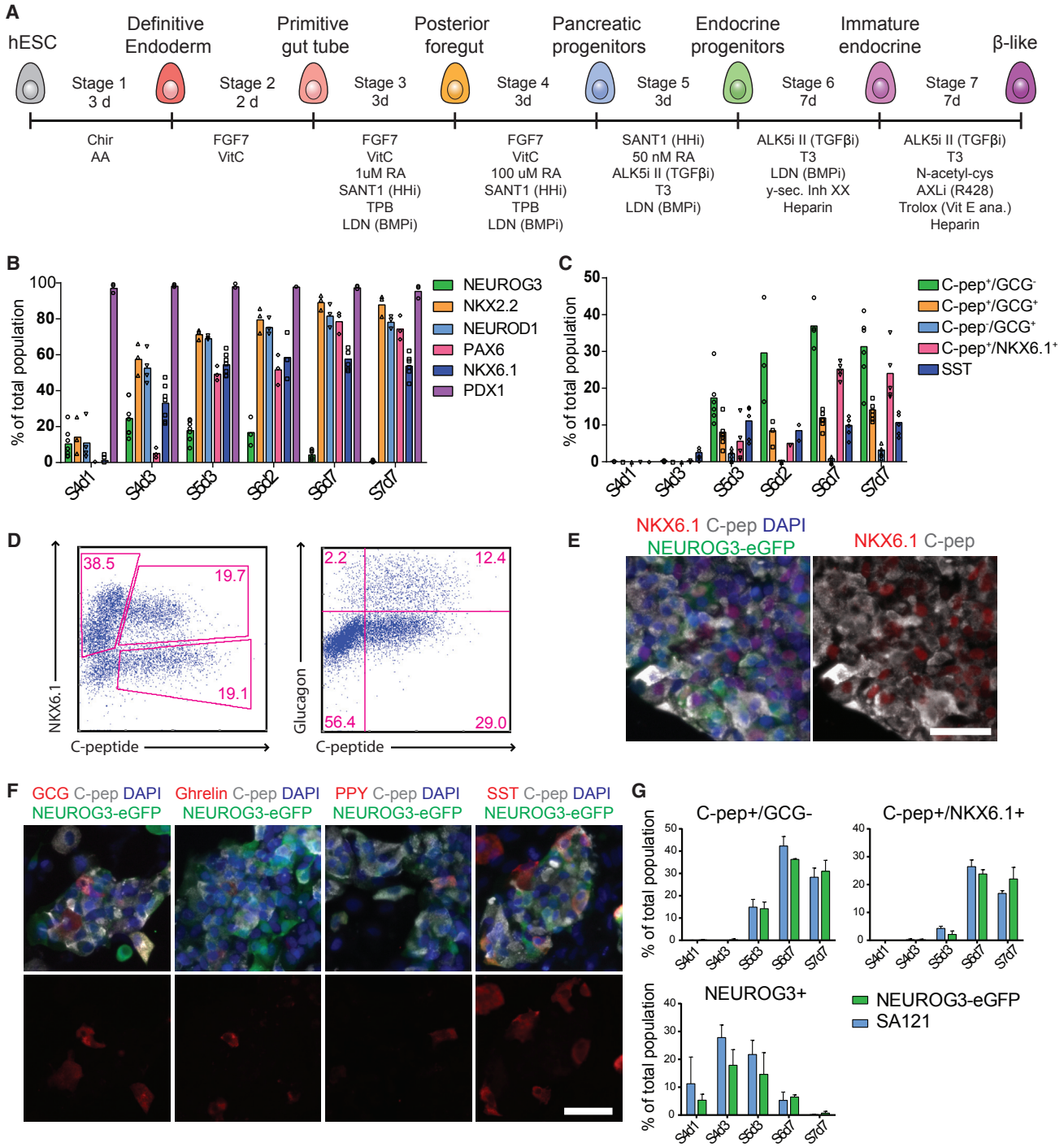


Figure 1. Generation of Pancreatic Endocrine Lineage Cells from hESCs

(A) Overview of 7-stage differentiation protocol.

(B and C) Flow-cytometry quantification of various transcription factors (B) and hormones (C) at six distinct stages of the differentiation protocol. Data are presented as individual biological replicates with error bars representing the mean (n = 3–7 except in C: S6d2 for C-peptide [C-pep]/NKX6.1 n = 1 and SST n = 2).

(D) Representative FACS plots for C-pep and NKX6.1 or C-pep and GCG in differentiated hESCs at S7d7.

(E and F) Immunofluorescence staining at S7d7 for EGFP, C-pep and NKX6.1 (E) or EGFP, C-pep, and the hormones GCG, ghrelin, PPY, or SST (F). Nuclei stained with DAPI. Scale bars, 50 μm.

(legend continued on next page)

RESULTS

In Vitro Generation of Pancreatic Endocrine Progenitors

To model human pancreatic endocrine development, we used an established 7-stage directed differentiation protocol (Rezania et al., 2014) with minor modifications (Figure 1A and Experimental Procedures) and a hESC line expressing EGFP under the control of the endogenous *NEUROG3* locus (NEUROG3-EGFP) (Löf-Öhlin et al., 2017). Similarly to several other lines, the NEUROG3-EGFP line differentiated efficiently to definitive endoderm and pancreatic progenitors, displayed robust endocrine induction marked by NEUROG3 protein expression during stages 4 and 5, and more mature endocrine cell differentiation at later stages (Figures S1A and S1B). At the final stage of the protocol we observed both C-peptide$^+$/glucagon$^-$ cells (β-like) and C-peptide$^+$/glucagon$^+$ cells (polyhormonal). Fifty-one percent of the C-peptide$^+$ cells co-expressed the β-cell marker NKX6.1 (Figures 1B–1E). We also observed some somatostatin$^+$ cells and rare cells expressing the hormones PPY or ghrelin (Figures 1C and 1F).

The longer half-life of GFP (Corish and Tyler-Smith, 1999) compared with NEUROG3 (Roark et al., 2012) leads to a decreased overlap with time and thus enables the capturing of cells that no longer express NEUROG3 but retain EGFP (Figures S1C–S1E). Conversely, the EGFP$^-$/NEUROG3$^+$ cells present at all stages (Figures S1D–S1F) are likely newly born NEUROG3$^+$ cells that have not yet accumulated enough EGFP protein to be detected, an interpretation confirmed by subsequent analysis below. Most EGFP$^+$ cells co-expressed NKX2.2, indicating progression into the endocrine lineage (Figure S1G).

Importantly, although the NEUROG3-EGFP hESC line used in this study is NEUROG3 haploinsufficient, it differentiated into endocrine progenitors and endocrine cells with efficiency similar to that of the parental cell line SA121 (Heins et al., 2004) (Figures 1G and S1H).

Together, these results demonstrate that we could recapitulate many aspects of the published protocol with the NEUROG3-EGFP line.

Transcriptional Profiling of Individual Endocrine Progenitors and Their Progeny

To explore heterogeneity within the endocrine progenitor population, we isolated NEUROG3-EGFP$^+$ and EGFP$^-$ cells at stage 4 day 1 (S4d1), when the first NEUROG3-EGFP$^+$ cells appeared, and at S4d3, S5d3, S6d2, S6d7, and S7d7, and analyzed them by single-cell qPCR (Figures 2A and S2A). In total, 546 cells were analyzed by single-cell qPCR for the expression of 86 genes, with 517 cells remaining for analysis after quality control assessment (Figure S2). The selection of genes encompassed key transcription factors involved in endocrine development and specification in the mouse, genes differentially expressed between mature human endocrine subtypes, and reporters for five major signaling pathways (Supplemental Experimental Procedures).

When visualizing the data using dimensionality reduction via t-distributed stochastic neighbor embedding (t-SNE) (Van Der Maaten and Hinton, 2008), the cells grouped in three major clusters (Figures 2B and S3A) classified as (1) pancreatic progenitors (*SOX9*$^+$), (2) early endocrine progenitors (*NEUROG3*$^+$), and (3) late endocrine progenitors and endocrine cells (*CHGA*$^+$) (Figures 2B–2D). Hierarchical clustering separated the cells into similar groups (Figure 2D). The pancreatic progenitors, which included most of the sorted EGFP$^-$ cells (regardless of stage), expressed pancreatic progenitor markers such as *PDX1*, *SOX9*, *ONECUT1*, *NOTCH1*, and *HES1* but did not express *NEUROG3* (Figures 2C and 2D). The second cluster consisted of mainly stage-4 cells and represented early endocrine progenitors expressing *NEUROG3* and *PAX4* but no hormones (Figures 2B and 2C). Nearly all EGFP$^+$ cells sorted at S4d1 contained *NEUROG3* transcript (97%), but the proportion decreased during differentiation (Figures S3C and S3D). Most of the cells that had *NEUROG3* transcript came from stages 4 and 5, confirming our flow-cytometry analysis (Figures 1B and S1D–S1F). The third cluster contained stage-5 and -6 *NEUROG3*$^+$ cells and maturing endocrine cells from stages 6 and 7. This cluster was characterized by high expression of *INS* and *CHGA* with most cells co-expressing *GCG* at varying levels (Figures 2B and 2C). Cells divided into two major lineages expressing either *NKX6.1* and *PDX1* or *ARX* and high levels of *GCG* (Figure 2C).

Distinct Pancreatic Progenitor Populations

A combination of t-SNE analysis, overlap of gene expression, and prior knowledge of gene expression hierarchy in mice enabled us to define several subtypes in each of the three clusters. Pancreatic progenitors transitioned from a state with no *NKX6.1* transcripts at S4d1 to most cells expressing *NKX6.1* as well as *MNX1* and other markers at S4d3 and S5d3 (Figure 3A). At S4d1, pancreatic

(G) FACS quantification of the percentage of cells expressing C-pep$^+$/GCG$^-$, C-pep$^+$/NKX6.1$^+$, and NEUROG3 throughout the differentiation of the NEUROG3-EGFP reporter cell line (heterozygous for NEUROG3; green bars) and the parental cell line SA121 (blue bars). Data are presented as mean ± SD (n = 3 biological replicates).
See also Figure S1.

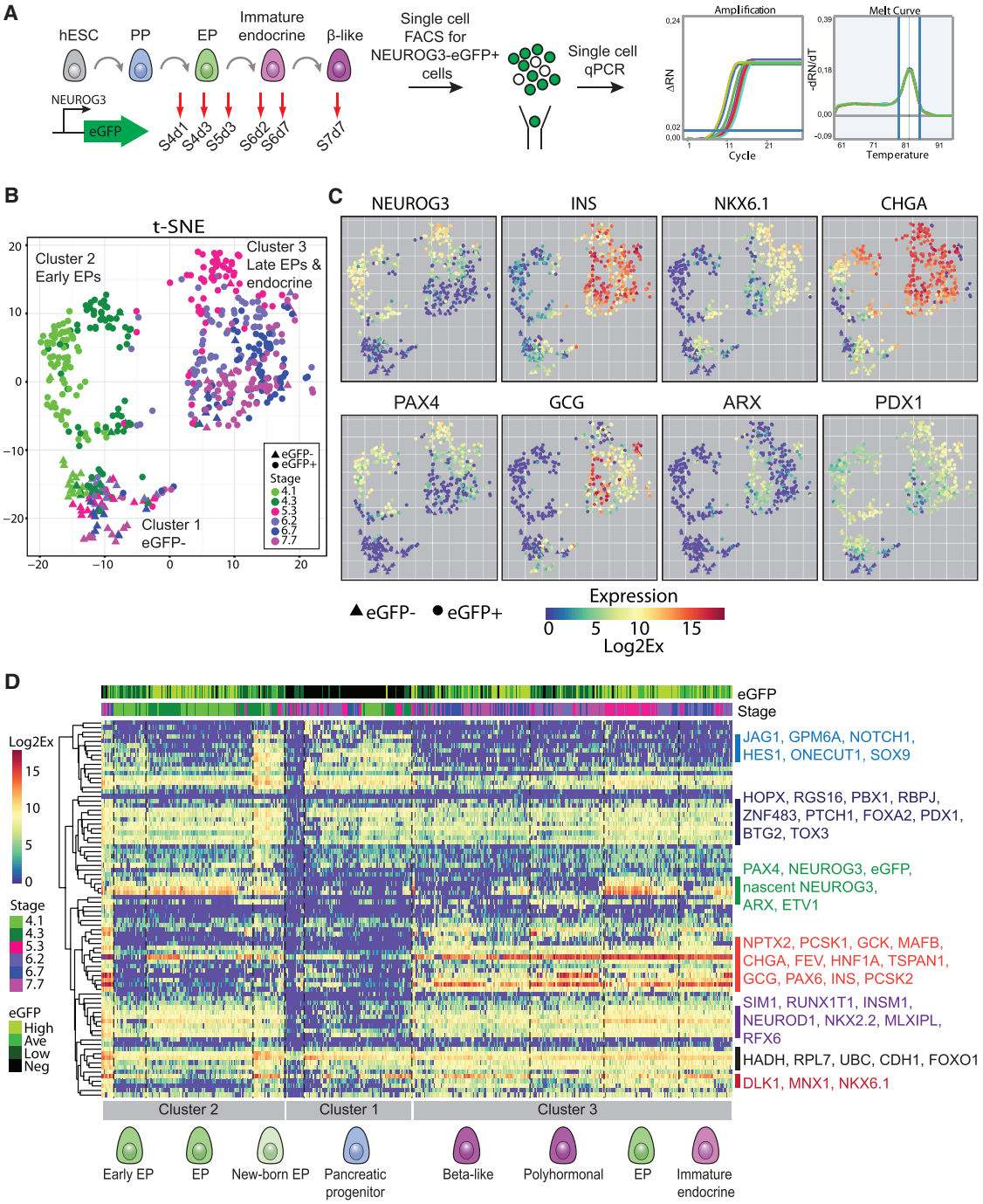


Figure 2. Single-Cell qPCR Analysis of Differentiated Endocrine Progenitors and Progeny

(A) Schematic showing the experimental setup. NEUROG3-EGFP hESCs were differentiated *in vitro* and NEUROG3-EGFP^{+} cells were sorted by FACS at the indicated time points for single-cell qPCR analysis.

(B) t-SNE analysis of 517 cells sorted from six time points during differentiation. Each dot represents a single cell and is colored according to differentiation stage with shape according to EGFP status (circle, EGFP^{+}; triangle, EGFP^{-}).

(C) t-SNE plots colored according to gene expression level of the indicated genes.

(D) Heatmap showing the expression level of 82 genes. Samples are displayed in columns and genes in rows. Data are clustered using average linkage method and distances calculated with Pearson correlation for samples and Euclidean distance for genes. The bars above the heatmap are color coded according to the level of EGFP protein expression in the individual sorted cell (top bar) and sample stage

(legend continued on next page)

progenitors expressed moderate levels of *RFX6*, *SIM1*, *INSM1*, and *HADH* (Figure 3A), which decreased at S4d3 and were lost at S5d3. Notably, these markers simultaneously became restricted to endocrine cells, as seen in mice for RFX6 (Soyer et al., 2010). *Sim1* and *Hadh* are likewise expressed during mouse pancreas development, but their cell-type-specific expression patterns have not been characterized (White et al., 2008).

EGFP$^-$ progenitors persisted beyond stage 5 but formed a heterogeneous group of cells, of which some appear to be misrouted toward different organ lineages. Some of these cells lacked *PDX1* expression and instead expressed the intestinal marker *CDX2* (Figure S3B). Expression of *ONECUT1* decreased with time, so that stage-7 EGFP$^-$ progenitors were largely negative, whereas *HES1* and *SOX9* decreased only in a subset of cells (Figure 3D). The simultaneous gain of *CHGA* and *INS* expression in the EGFP$^-$ population at S7d7 indicates that virtually all cells in the culture at this stage have initiated some level of endocrine differentiation (Figure 3D). However, the low levels of *CHGA* and *INS* and the rare expression of *NKX2.2*, *NEUROD1*, and *PAX6* (Figure 3D) suggest that these cells have not committed fully to the endocrine lineage, or are perhaps reverting.

Heterogeneity within the *NEUROG3*$^+$ Endocrine Progenitor Population

Many genes were differentially expressed in endocrine progenitors between stage 4 and later stages, but we also observed contemporary subpopulations, suggesting that cells do not differentiate at the same pace or indicating the existence of lineage bifurcations. At S4d1, a population of *NEUROG3*$^+$ cells co-expressed the pancreatic progenitor markers *SOX9*, *HES1*, and *ONECUT1* and is expected to represent the first step of endocrine commitment (Figure 3A). These cells did not express *NKX6.1*, in agreement with the absence of *NKX6.1* in pancreatic progenitor cells at S4d1 (Figure 3B). In addition, these cells expressed *NKX2.2* and *RUNX1T1*, which appear to be the first NEUROG3 targets. At the same stage, a more mature population was identified with *PAX4*, *MAFB*, *NEUROD1*, low *CHGA*, and *GCK* additionally induced while the progenitor markers *SOX9*, *HES1*, and *ONECUT1* were downregulated.

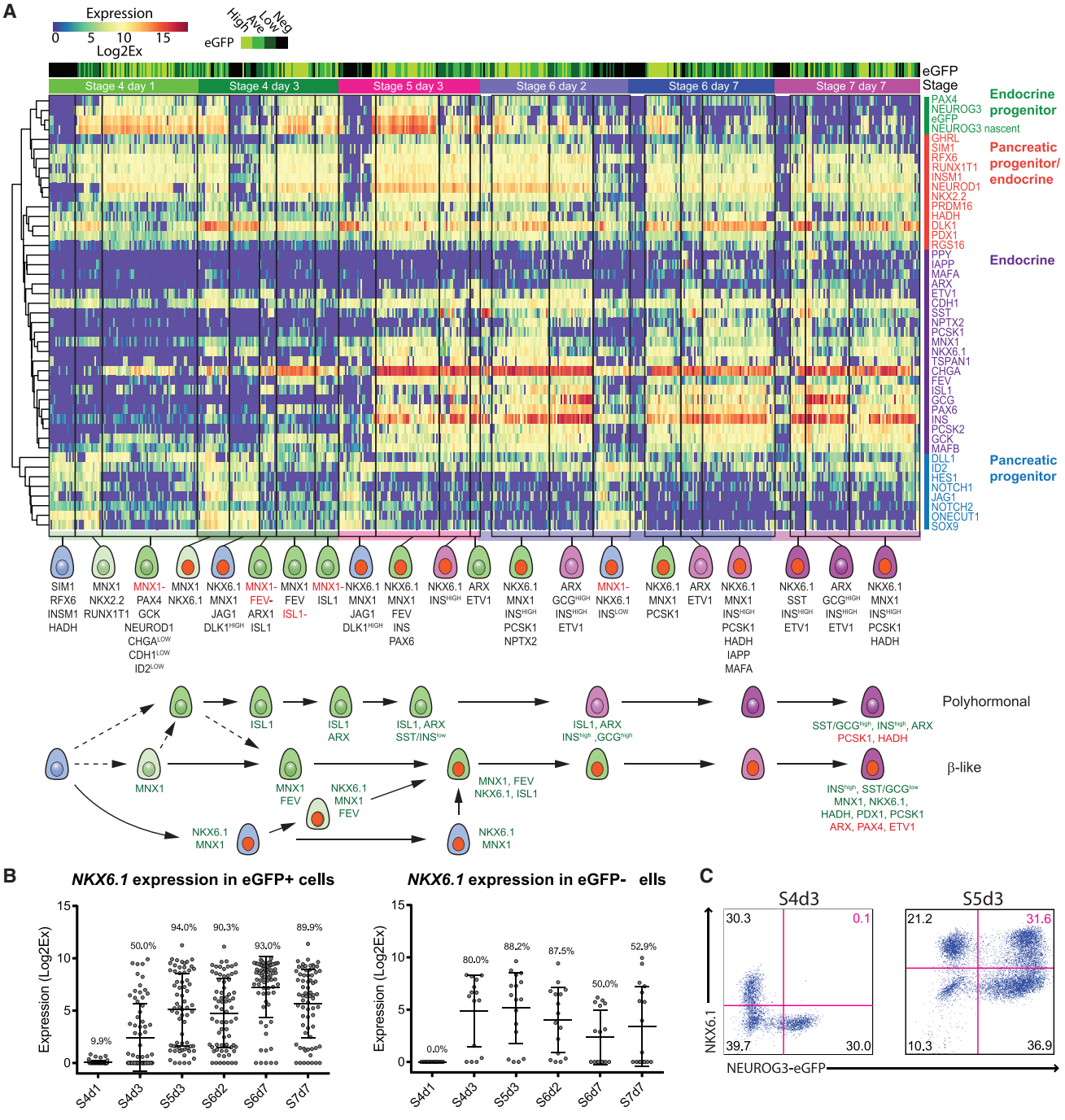
While most EGFP$^-$ pancreatic progenitors expressed *NKX6.1* at S4d3, only a subset of EGFP$^+$ cells had *NKX6.1* transcripts (Figure 3B) and none expressed NKX6.1 protein at this stage (Figure 3C), indicating that the early NEUROG3$^+$ cells do not derive from an NKX6.1$^+$ population. The presence of *NKX6.1* transcripts in some EGFP$^+$ cells suggests the initiation of NKX6.1 expression after endocrine commitment. At S4d3, a *NEUROG3*$^+$ population co-expressed pancreatic progenitor markers, suggesting that these cells entered the differentiation pipeline at this stage. These cells also expressed *NKX6.1*, unlike their counterpart at S4d1, as well as *PAX4*, *MAFB*, *NEUROD1*, *CHGA*, *GCK*, and low levels of *PAX6* and *FEV*. They likely originate from *NKX6.1*$^+$/*MNX1*$^+$ pancreatic progenitors. Two other more mature populations of endocrine progenitors expressed *NEUROG3* but no pancreatic progenitor markers. None expressed *NKX6.1*, and they thus likely derive from the *NKX6.1*$^-$ progenitors present earlier. While the most abundant group expressed *PAX6*, *FEV*, and low *MNX1*, the other expressed *ISL1* and *ETV1* (Figure 3A). A population expressed similar markers to the second but more NEUROG3 targets such as *PAX6* and *ARX*, as well as low levels of insulin, while the levels of *NEUROG3* and *PAX4* were reduced. This suggests a maturation step further representing the first transition to hormone expression. Taken together, these findings suggest a divergence of differentiation trajectories during stage 4.

The expression level of *NEUROG3* (both mature and nascent mRNA transcript) and *EGFP* peaked in endocrine progenitors of stage 5 (Figure 3A). Endocrine progenitors and progeny continued to mature at S5d3, increasing *CHGA* and expressing the hormone *INS* as well as *MNX1*, *PAX6*, *FEV*, *ISL1*, *PCSK2*, and *TSPAN1*. *NKX6.1* expression partitioned two different *NEUROG3*$^+$ populations: a large population characterized by *NKX6.1*$^+$, *HADH*$^+$, and low levels of *PCSK1* expression while having low or no expression of *SST* and *GCG* hormone transcripts, and a smaller group of *NKX6.1*$^-$ cells that had higher levels of *GCG* or *SST*, co-expressed *ARX*, but not *HADH* (Figure 3A). Additionally, a population of immature endocrine cells was present at stage 5 that had turned off *NEUROG3*, while some had also lost *PAX4* expression. Most of these cells expressed *NKX6.1* and *PCSK1*. Endocrine progenitors were also found at stage 6, although they had lower levels of *NEUROG3* and *PAX4* than those found at stages 4 and 5. These cells were similar to the *NKX6.1*$^+$ endocrine progenitors at stage 5, but additionally expressed *ISL1*, more robust *PCSK1*, and moderate levels of *GCG* (Figure 3A).

Co-transcription of Multiple Hormones during *In Vitro* Differentiation

We observed a substantial co-expression of multiple hormone transcripts during the *in vitro* differentiation of

(bottom bar). Gray bars below the heatmap indicate the corresponding clusters in the t-SNE in (B). Major cell populations are indicated below the heatmap. The cells originate from three independent biological replicates of the entire differentiation, carried out on different days, in similar proportions, with no batch effect.
See also Figures S2 and S3.



(legend on next page)

hESCs. As early as S4d1, low levels of ghrelin transcripts were detected in *NEUROG3*$^+$ cells but not in pancreatic progenitors (Figure 3A). The next hormone transcripts appearing were *SST* and *INS* at S4d3, followed by *GCG*. Ghrelin transcripts were expressed in a surprisingly large proportion of endocrine committed cells until stage 7, when expression was lost in most cells. *PPY* transcripts were detected in a few cells from S5d3 and beyond (Figure 3A).

Most *INS*$^+$ cells co-expressed *GCG* and/or *SST* transcripts at varying levels from stage 6 onward (Figures 2C and 3A), although this transcriptional co-expression does not lead to detectable hormone production in most cells (Figures 1C and 1D). We speculate that cells with high *GCG* transcript levels and *ARX* expression are also bihormonal at the protein level, as seen during human development (Jeon et al., 2009; Riedel et al., 2012), whereas cells with *NKX6.1* expression and often lower levels of *GCG* will not express detectable levels of glucagon protein. Alternatively, a post-transcriptional regulation may occur. Cells did not seem to mature much during stage 7; on the contrary, many genes important for β-cell identity were expressed at higher level at S6d7 compared with S7d7, including *NKX6.1* and *PCSK1*. At S6d7 many endocrine cells displayed a β-cell-like profile, expressing *INS*, *NKX6.1 MNX1*, *PCSK1*, *PCSK2*, *NPTX2*, *HADH*, and some *MAFA* and *IAPP*. Such a population was also found at S7d7, although smaller, whereas the largest endocrine population at S7d7 had decreased levels of *NPTX2* and *HADH*, and only sporadic cells expressed *MAFA* and *IAPP* (Figure 3A).

Early Endocrine Progenitors Initiate NKX6.1 Expression Downstream of NEUROG3

Our single-cell qPCR analysis suggested a potential branching point during differentiation of *NEUROG3*$^+$ endocrine progenitors between stages 4 and 5 characterized by the acquisition of *NKX6.1* transcript. As early endocrine induction in NKX6.1$^-$ cells in hPSC-derived cultures has been proposed to give rise to INS$^+$/GCG$^+$ polyhormonal cells *in vitro* and α-like cells upon engraftment in mice (Kelly et al., 2011; Rezania et al., 2013), we sought to investigate the potential of endocrine progenitors formed during stage 4 and those born during stage 5 or thereafter. No cells expressed NKX6.1 protein at S4d1 (Figure 4A) and the transcript was restricted to 10% NEUROG3-EGFP$^+$ cells (Figure 3B). NKX6.1 was restricted to some NEUROG3-EGFP$^-$ cells at S4d3 but was excluded from NEUROG3$^+$ and NEUROG3-EGFP$^+$ cells, although many had *NKX6.1* transcripts (Figure 4A). NKX6.1 and NEUROG3-EGFP were co-expressed by 31% ± 8% of the total cell population at S5d3. As most EGFP$^+$ cells at S4d3 expressed NKX2.2 protein (Figure S1G), this suggests that they have initiated a transcriptional program downstream of NEUROG3 and are committed to the endocrine lineage. These findings suggest that NKX6.1 expression can be acquired downstream of NEUROG3 expression and endocrine commitment.

To explore this further, we isolated NEUROG3-EGFP$^+$ and EGFP$^-$ cells by fluorescence-activated cell sorting (FACS) at S4d3, and cultured the two populations further in stage-5 media to test their fate. Due to the limited number of cells we could retrieve from sorting and to have sufficient numbers of cells for later analysis, we co-cultured the sorted cells with S4d3 cells derived from the SA121 hESC line differentiated in parallel (Figure 4B and Supplemental Experimental Procedures). After 3 days of culture, cells were fixed and stained for NKX6.1. Immunofluorescence staining and flow cytometry revealed the presence of NEUROG3-EGFP$^+$ cells in the cultures with sorted S4d3 NEUROG3-EGFP$^+$ cells, showing that cells can retain EGFP reporter signal for at least 3 days (Figures 4C and 4E). NEUROG3-EGFP$^+$ cells were similarly detected in the

Figure 3. Distinct Cell Populations Identified at Each Stage of Differentiation

(A) Heatmap displaying the gene expression level in 517 single cells within each of the six stages of differentiation for 45 selected genes. Samples are displayed in columns and genes in rows. Data are clustered using average linkage method and distances calculated with Pearson correlation for samples (per stage) and Euclidean distance for genes. The bars above the heatmap are color coded according to the level of EGFP protein expression in the individual sorted cell (top bar) and sample stage (bottom bar). Cell populations found at each stage are indicated below the heatmap with their associated key transcription factors and other markers. Blue cells represent pancreatic progenitors; light-green cells represent newly formed endocrine progenitors; green cells represent endocrine progenitors; light-purple cells represent immature endocrine cells; dark-purple cells represent endocrine cells. Red nuclei indicate *NKX6.1*$^+$ expression. The proposed lineage relationships between the cell types at different time points are shown below.

(B) Scatterplot showing the level of *NKX6.1* transcript in NEUROG3-EGFP$^{+/-}$ cells from each stage. Horizontal line represents mean expression and error bars represent SD (number of cells is shown in Figure S3C). Numbers above each plot indicate the percentage of NEUROG3-EGFP$^{+/-}$ cells with *NKX6.1* expression >0.

(C) Representative flow-cytometry plot showing expression of NEUROG3-EGFP and NKX6.1 at S4d3 and S5d3. Numbers in each quadrant indicate percentage of cells.

(D) Violin plots for the distribution of gene expression per stage of the indicated genes in the NEUROG3-EGFP$^+$ population versus the NEUROG3-EGFP$^-$ population. The cells originate from three independent biological replicates of the entire differentiation, carried out on different days, in similar proportions, with no batch effect.

See also Figure S3.

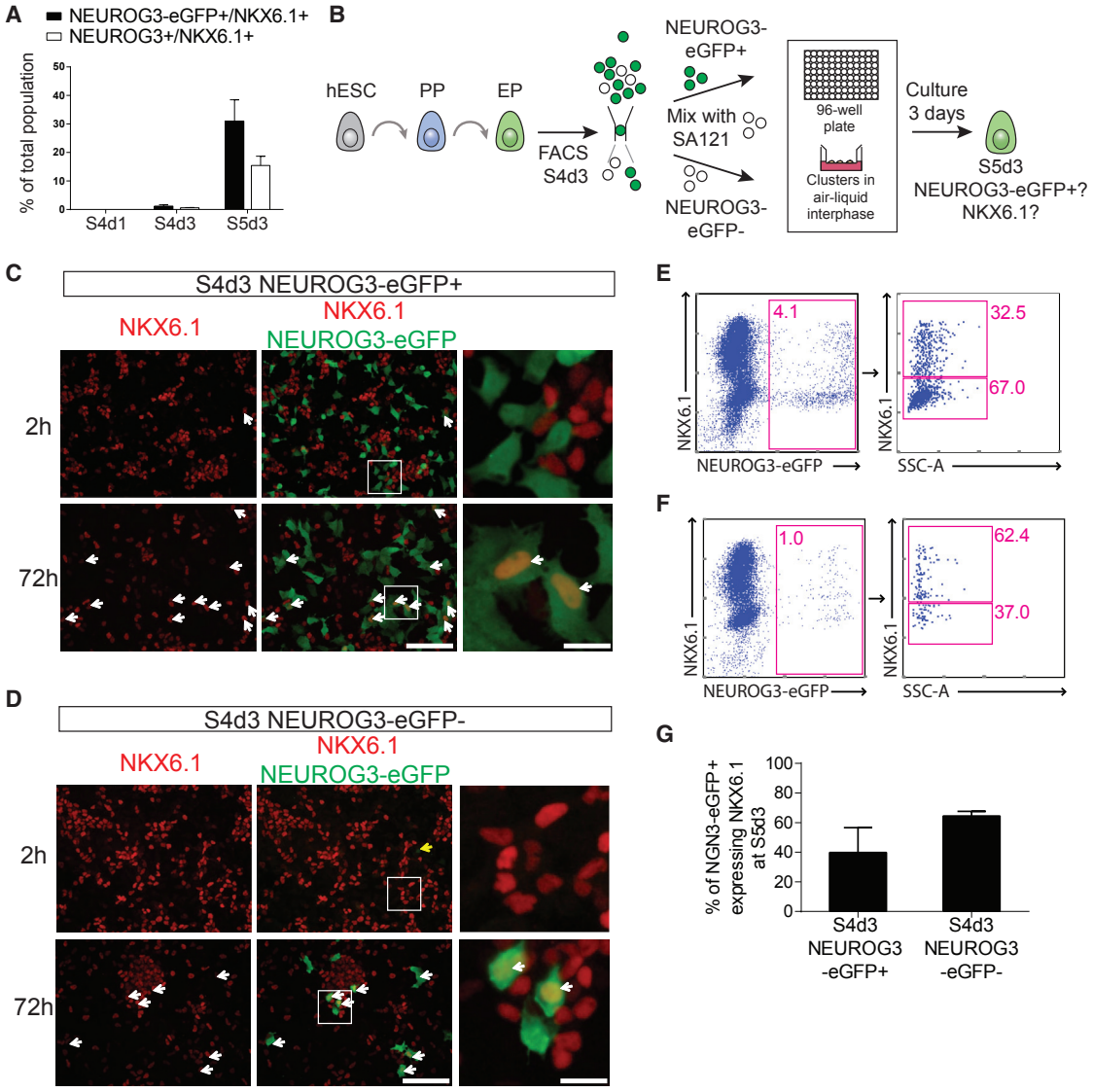


Figure 4. NKX6.1 Is Induced Downstream of NEUROG3

(A) Flow-cytometry analysis of co-expression of NEUROG3-EGFP and NKX6.1 at S4d1, S4d3, and S5d3 (black bars, mean ± SD, n = 3) or NEUROG3 and NKX6.1 (white bars, mean ± SD, n = 2).

(B) Schematic showing the experimental setup. NEUROG3-EGFP$^{+/-}$ cells were sorted at S4d3, mixed with SA121 hESC differentiated in parallel, and cultured in either 2D or 3D to S5d3. Cells were then analyzed for EGFP and NKX6.1 expression.

(C and D) Immunofluorescence staining for EGFP and NKX6.1 of S4d3 sorted NEUROG3-EGFP^{+} cells (C) and NEUROG3-EGFP^{-} cells (D) 2 hr and 72 hr after plating in 2D. White arrows indicate NKX6.1^{+}/EGFP^{+} cells. The yellow arrow marks a single weak EGFP^{+} cell. Scale bars, 100 μm; scale bars in enlargements, 20 μm.

(E and F) Representative FACS plots for NKX6.1 and NEUROG3-EGFP for S4d3 sorted NEUROG3-EGFP^{+} cells (E) and NEUROG3-EGFP^{-} cells (F) cultured in 3D for 72 hr (to S5d3).

(G) Flow-cytometry quantification of NEUROG3-EGFP$^{+/-}$ cells sorted at S4d3 and analyzed at S5d3 after 72 hr of culturing in 3D for co-expression of NEUROG3-EGFP and NKX6.1. Data represent mean ± SEM (n = 3 for S4d3 NEUROG3-EGFP^{+}-derived cells and n = 2 for S4d3 NEUROG3-EGFP^{-}-derived cells). n refers to biological replicates of the entire differentiation, carried out on different days.

See also Figure S4.

cultures with sorted S4d3 NEUROG3-EGFP^{-} cells, demonstrating that new NEUROG3^{+} cells arise during stage 5 (Figures 4D and 4F). Among these, 65% ± 3% co-expressed NKX6.1, as revealed by flow-cytometry analysis of 3D cultures (Figure 4G). While S4d3 NEUROG3-EGFP^{+} cells had minimal co-expression of NKX6.1, many of the sorted

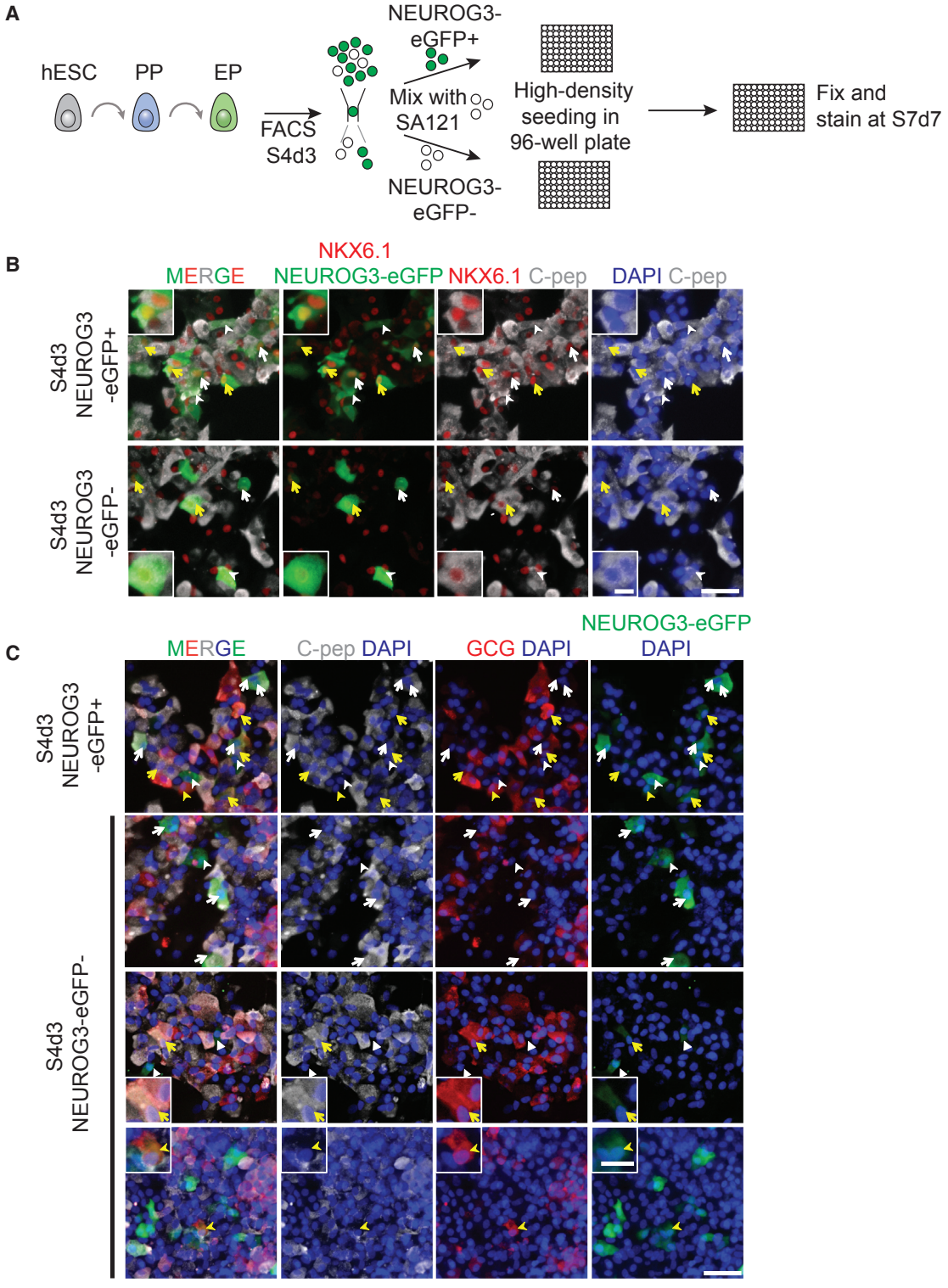


Figure 5. Early- and Late-Born Endocrine Progenitors Both Give Rise to NKX6.1^{+}/C-pep^{+} Cells at S7d7

(A) Schematic showing the experimental setup. NEUROG3-EGFP$^{+/-}$ cells were sorted at S4d3, mixed with SA121 hESC differentiated in parallel, and cultured in 2D to S7d7. Cells were then fixed and stained at S7d7 for EGFP, NKX6.1, C-pep, and GCG.

(legend continued on next page)

S4d3 EGFP$^+$ cells expressed NKX6.1 when analyzed at S5d3 by flow cytometry (40% ± 17%) (Figures 4E and 4G) or immunocytochemistry (Figures 4C, 4D, and S4A). We thus conclude that NEUROG3-EGFP$^+$ cells are born during stages 4 and 5, and our sorting experiment further demonstrates that NKX6.1 can be acquired in cells after they have initiated NEUROG3 expression.

Early- and Late-Born Endocrine Progenitors Have the Same Developmental Potential

Next, we explored whether acquiring NKX6.1 downstream of NEUROG3 would enable the generation of β-like cells. We co-cultured EGFP$^+$ or EGFP$^-$ cells sorted at S4d3 with SA121 to S7d7 and found that both early- and late-formed NEUROG3-EGFP$^+$ cells gave rise to EGFP$^+$ cells that co-expressed C-peptide and NKX6.1 at S7d7 (Figures 5A and 5B). Additionally, both populations gave rise to NKX6.1$^+$ progeny that did not express C-peptide as well as C-peptide$^+$ cells devoid of NKX6.1. Likewise, we detected EGFP expression in C-peptide$^+$/glucagon$^+$ cells, C-peptide$^+$/glucagon$^-$, cells and very rare C-peptide$^-$/glucagon$^+$ cells derived from both populations of endocrine progenitors (Figure 5C). Early- and late-born endocrine progenitors thus have a similar potential to generate C-peptide$^+$/NKX6.1$^+$ or C-peptide$^+$/GCG$^+$ cells *in vitro*.

We additionally co-cultured EGFP$^+$ or EGFP$^-$ cells sorted at S4d3 with likewise differentiated SA121 cells in 3D at the air-liquid interface and FACS isolated NEUROG3-EGFP$^+$ cells at S7d7 for single-cell qPCR (Figure 6A). Among the 41 S4d3 EGFP$^+$-derived cells and 38 S4d3 EGFP$^-$-derived cells that passed quality assessment, ANOVA analysis comparing the two populations identified seven slightly differentially expressed genes, including *INS*, *GCG*, and *HADH* (Figures 6B and S4D), but this did not lead to a segregation of the two populations in t-SNE (Figure 6C) and hierarchical clustering (Figure 6D). Hierarchical clustering using selected genes important for β-cell identity separated the data into a cluster characterized by co-expression of *INS* and *GCG* transcripts, while the other had low or no *GCG* transcripts in addition to expression of *INS*, *NKX6.1*, *PDX1*, *PCSK1*, and *HADH* (Figure 6D). Importantly, cells derived from both early- and late-born endocrine progenitors contributed to both populations.

In summary, these experiments reveal that NEUROG3-EGFP cells born during stage 4 have the potential to give rise to the same cell types as those born during or after stage 5. NEUROG3-EGFP$^+$ cells devoid of NKX6.1 protein can acquire it after endocrine commitment and become *INS*$^+$/*GCG*$^-$-low/negative cells that further express β-cell markers such as *PCSK1*, *IAPP*, and *HADH*.

To further investigate the profile of endocrine progenitors giving rise to INS$^+$/NKX6.1$^+$ cells versus GCG$^+$/INS$^+$/NKX6.1$^-$ cells, we compared early- and late-born NEUROG3-EGFP$^+$ cells generated with the described 7-stage protocol ("new") with corresponding cells produced using an earlier developed 4-stage protocol ("old"), which gives rise to mostly polyhormonal GCG$^+$/INS$^+$/NKX6.1$^-$ cells *in vitro* (Rezania et al., 2012) (Figure S5). Cells at the late stage of the old protocol had lost NEUROG3 expression with most cells co-expressing *INS*, *GCG*, and *ARX*, whereas *NKX6.1*, *MNX1*, *HNF1A*, and *PCSK2* transcripts were largely confined to the new protocol (Figures S5B and S5C). Interestingly, most early endocrine progenitors of the old protocol expressed *ISL1* but not *FEV*, whereas only a smaller subset of *NEUROG3*$^+$ cells from the new protocol expressed *ISL1*. At the late stage, cells of the old protocol all expressed *ISL1* and lower levels of *FEV*, whereas cells of the new protocol had *FEV* but not all had *ISL1*. This is coherent with our earlier observations suggesting that *FEV* expression before *ISL1* in endocrine progenitors marks cells on the β-cell path, while *ISL1* onset before *FEV* appears to predict the polyhormonal *INS*$^+$/*GCG*$^+$/*ARX*$^+$/*NKX6.1*$^-$ profile seen at the end of the old protocol.

Transcriptional Similarities and Differences between *In-Vitro*-Generated β-like Cells and Adult Human β Cells

To assess the similarity between cells produced *in vitro* to mature islet cells, we performed single-cell qPCR on 165 human islet cells, encompassing α (107), β (18), PP (8), δ (2), cells and a group of non-endocrine cells (30), of which 22 had a ductal gene expression profile and 8 had a mesenchymal profile (Figure S6). Comparison between the different cells is presented in Figure S6 and the main findings are reported in Supplemental Information.

We next compared the human islet single-cell data with that of hESC-derived cells. T-SNE analysis showed that the non-endocrine cells from human islets clustered closely to the pancreatic progenitor cells, the β cells were in proximity to the stage-6 and -7 cells, and the α and PP cells formed a cluster on their own at equal distance to all other cells (Figure 7A). We then compared the gene expression profile of human α and β cells with hESC-derived *INS*$^+$ cells divided into four groups based on hierarchical clustering

(B and C) Immunofluorescence staining at S7d7 of S4d3 sorted NEUROG3-EGFP$^+$ cells and NEUROG3-EGFP$^-$ cells cultured in 2D along with SA121 cells. (B) Staining for EGFP, NKX6.1, C-pep, and DAPI. White arrows, EGFP$^+$/C-pep$^-$/NKX6.1$^+$; white arrowheads, EGFP$^+$/C-pep$^+$/NKX6.1$^-$; yellow arrows, EGFP$^+$/C-pep$^+$/NKX6.1$^+$. (C) Staining for EGFP, C-pep, GCG, and DAPI. White arrows, EGFP$^+$/C-pep$^+$/GCG$^-$; white arrowheads, EGFP$^+$/C-pep$^-$/GCG$^-$; yellow arrows, EGFP$^+$/C-pep$^+$/GCG$^+$; yellow arrowheads, EGFP$^+$/C-pep$^-$/GCG$^+$. Scale bars, 50 μm; scale bar in insets in (B), 12.5 μm.

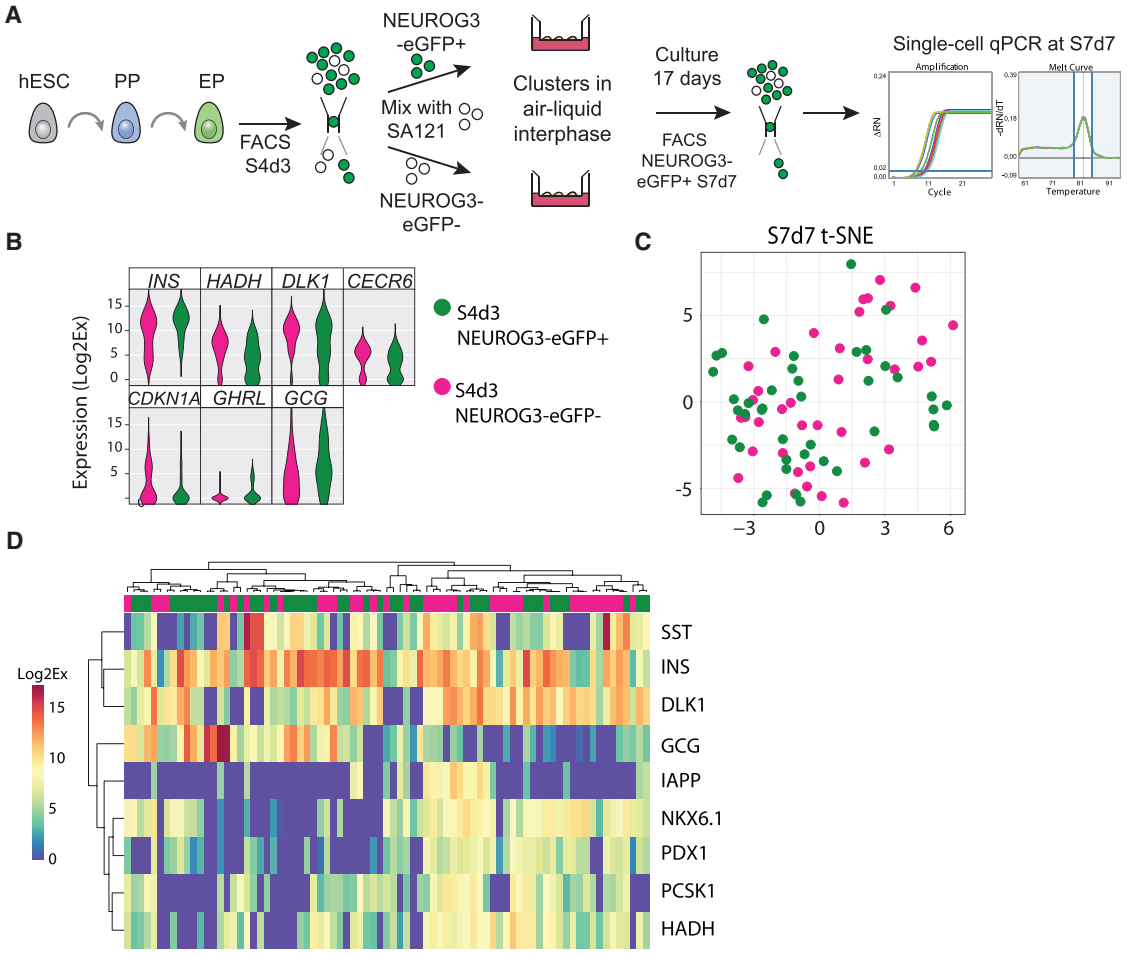


Figure 6. Potential of hESC-Derived Endocrine Progenitors

(A) Schematic showing the experimental setup. NEUROG3-EGFP$^{+/-}$ cells were sorted at S4d3, mixed with SA121 hESC differentiated in parallel, and cultured in 3D to S7d7. Cells were then sorted by FACS at S7d7, and EGFP^{+} cells collected for single-cell qPCR analysis.

(B) Violin plot showing the distribution of gene expression level of seven genes differentially expressed based on ANOVA analysis between single EGFP^{+} cells at S7d7 derived from early- or late-born endocrine progenitors.

(C) t-SNE analysis of EGFP^{+} progeny at S7d7 from early- or late-born endocrine progenitors.

(D) Heatmap showing the expression level of selected genes in EGFP^{+} progeny at S7d7. The cells originate from three independent biological replicates of the entire differentiation, carried out on different days, in similar proportions, with no batch effect.

See also Figure S5.

(Figure S6F). One group represented polyhormonal cells, while cells with a more β-cell-like profile were divided into three groups based on maturation stage: β-like (*NEUROG3*$^{-}$/*EGFP*$^{-}$), immature β-like (*NEUROG3*/*EGFP* low), and early immature β-like (*NEUROG3*$^{+}$/*EGFP*$^{+}$) (Figures 7B and S6F). Analyzing the expression of genes enriched in human β cells, we found equal levels of *PTCH1*, *RUNX1T1*, *TSPAN1*, *RGS16*, *DLK1*, *PDX1*, *MNX1*, and *NKX6.1* between human β cells and the three groups of β-like cells (Figure 7C). *INS* expression was highest in β cells, followed by β-like cells, polyhormonal cells, and finally the two groups of immature β-like cells. *PCSK1* was expressed in β cells and all β-like and immature β-like cells, whereas just a subset of the early immature β-like cells had *PCSK1* (Figure 7B). These findings show that a subset of the *in-vitro*-generated cells share important transcriptional identity with primary human β cells.

Nevertheless, we also identified genes differentially expressed between β cells and hESC-derived cells. These included *GPM6A*, *MAFA*, *HDAC9*, *IAPP*, and *HADH*, which were expressed in all β cells at much higher level compared with even the most mature β-like cells (Figure 7D), showing a lack of maturity of *in-vitro*-produced cells. On the contrary, *KRT19* was expressed at higher level in hESC-derived

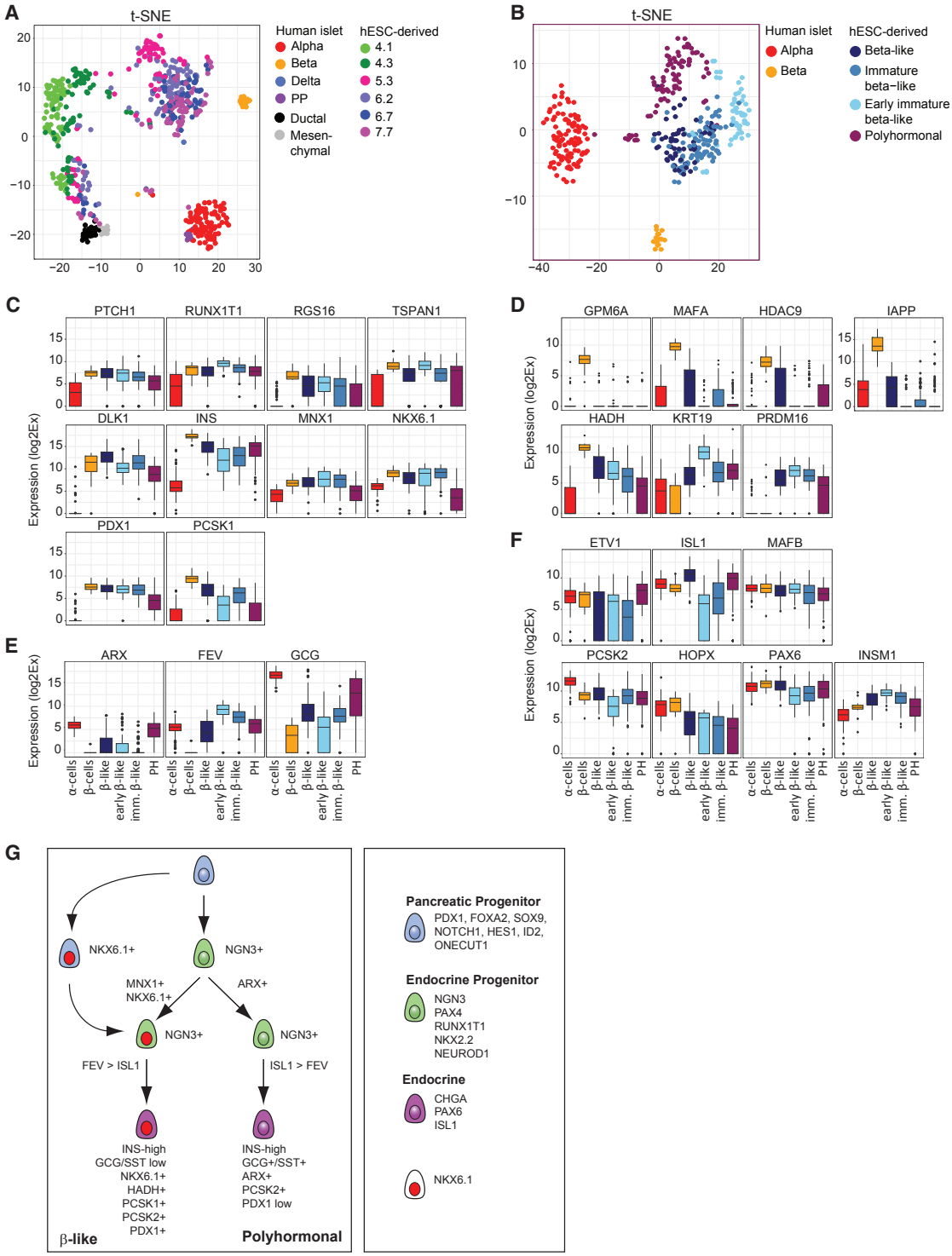


Figure 7. Comparison of Human Adult β and α Cells with hESC-Derived *INS*$^+$ Cells

(A) t-SNE plot of hESC-derived cells colored by stage and human islet cells colored by cell type.

(B) t-SNE plot of human α and β cells along with hESC-derived *INS*$^+$ cells categorized as polyhormonal, β-like, immature β-like, and early immature β-like.

(legend continued on next page)

cells compared with human α and β cells, whereas *PRDM16* was exclusive to the hESC-derived cells, indicating that these markers represent immaturity. The α-cell-enriched genes *ARX* and *FEV* were expressed at equal levels in α cells and polyhormonal cells, whereas *FEV* was also expressed in β-like cells, with higher expression in the immature subgroups (Figure 7E). *GCG* was expressed at very high level in all α cells, more heterogeneously in polyhormonal cells, lower in β-like cells, and lowest in β cells (Figure 7E).

Finally, we explored the expression pattern of genes equally expressed in human α and β cells. *MAFB*, *PAX6*, and *INSM1* were expressed at similar level in α, β, and hESC-derived cells (Figure 7F). *ETV1* was expressed in all α, β, and polyhormonal cells, but only in a subset of β-like cells. *PCSK2* was expressed at similar level in all cells, though slightly higher in α cells. *HOPX* was expressed in α and β cells and most hESC-derived β-like cells, but only in some immature β-like cells and polyhormonal cells.

DISCUSSION

In this study, we present a comprehensive characterization of hESC-derived human pancreatic endocrine progenitors and their progeny by use of single-cell qPCR and compare their transcriptional profile with that of primary human islet cells. Our single-cell analysis identifies subpopulations among pancreatic progenitors, endocrine progenitors, and endocrine cells during directed differentiation of hESCs toward the pancreatic endocrine lineage. These observations enable us to delineate progression steps along the differentiation path and prospective differentiation trajectories. We find that *MNX1* and *NKX6.1* distinguish populations of pancreatic progenitor cells at S4d1 and S4d3, and newly identify the expression of *SIM1*, *RFX6*, *INSM1*, and *HADH* in early human pancreatic progenitors. We find that hormone expression is already initiated in the *NEUROG3*$^+$ cells. Signs of lineage bifurcations and convergence among endocrine progenitors were observed as early as S4d3, with distinct gene expression signatures between endocrine progenitor subpopulations based on *MNX1*, *NKX6.1*, *ARX*, *FEV*, and *ISL1*. These bifurcations lead to different populations of endocrine cells at stages 6 and 7, manifesting with either a polyhormonal signature (high *INS* and *GCG*, *ARX*$^+$) or a β-cell signature (high *INS* and *PCSK1*$^+$, *NKX6.1*$^+$). Furthermore, our findings uncover a very long period of continuous *NEUROG3* induction and co-expression of multiple hormonal transcripts during *in vitro* hESC differentiation.

Endocrine Lineage Trajectories Suggest Multiple Routes to β-like Cells

The marker overlap between populations and information from mouse pancreas development leads us to propose a general endocrine maturation sequence where intermediate, possibly plastic states are identified. While newly born endocrine progenitors expressing *NEUROG3* retain pancreatic progenitor markers (*SOX9*, *ONECUT1*, *HES1*, *NOTCH1/2*), the first endocrine markers initiated downstream of *NEUROG3* are *RUNX1T1* and *NKX2.2*, followed by *PAX4* and *NEUROD1* and subsequently *CHGA*, *GCK*, and the prohormone convertases *PCSK1/2*. These transcription factors are NEUROG3 targets in mice and may be activated with different kinetics (Benitez et al., 2014).

In addition, we provide evidence for early lineage bifurcation at S4d1 or S4d3. This bifurcation may already occur at S4d1 between newly formed endocrine progenitors (still retaining most pancreatic progenitor markers) expressing *MNX1*, but not *PAX4* or *NEUROD1* and a large population of *PAX4*$^+$, *NEUROD1*$^+$, *MNX1*$^-$ cells (Figures 3A and S7). These two populations may form independently or may be steps on the same path, as suggested by the maturity profile, which would imply the loss of *MNX1* as the first population matures into the latter. The S4d1 *MNX1*$^+$ population will likely give rise to the more mature *MNX1*$^+$ endocrine progenitors found at S4d3, which express markers of progression along the endocrine path (*PAX4*, *NEUROD1*, and *PAX6*) as well as a new marker, *FEV* (Figures 3A and S7). Likewise, the S4d1 *MNX1*$^-$ endocrine population is proposed to give rise to the *MNX1*$^-$ population at S4d3 that expresses *ISL1* but not *FEV*, and progresses further to *ARX*$^+$ cells, as *NEUROG3* and *PAX4* transcript levels are downregulated. We cannot rule out that the S4d1 *MNX1*$^-$ population might also contribute to the β-cell fated lineage, if cells can acquire *MNX1* at this stage (dashed arrow in Figure S7). In addition to these endocrine progenitor populations found at S4d3, pancreatic progenitors at this stage are expressing *MNX1* and *NKX6.1*, and give rise to the population of newly formed endocrine progenitors

(C–F) Box-and-whisker plots showing gene expression in α-, β-, and hESC-derived INS$^+$ cells for genes upregulated in β cells and hESC-derived β-like cells (C), genes differentially expressed in β-cell and hESC-derived β-like cells (D), α-cell-specific genes (E), and genes expressed at equal levels in both α and β cells (F). The boxes represent the 25th (Q1) and 75th (Q3) percentiles, and whiskers are at 1.5 times interquartile range from Q1 and Q3.

(G) Model of *in vitro* endocrine differentiation toward the two major lineages of β-like versus polyhormonal cells. The hES-derived cells originate from three independent biological replicates of the entire differentiation, carried out on different days, in similar proportions. The islet cells originate from a single individual with high-quality expression.

See also Figures S6 and S7.

present at S4d3, which are *MNX1*$^+$ and *NKX6.1*$^+$. These cells also express *FEV* but not *ISL1*.

In summary, we suggest that *MNX1* expression in early *NKX6.1*$^-$ endocrine progenitors mark cells that will become *NKX6.1*$^+$, while *FEV* expression preceding *ISL1* initiation is also marking this fate choice (Figure 7G). We propose that this differentiation path is feeding into the lineage fated to a β-like cell type in parallel with endocrine cells derived from *NKX6.1*$^+$/*MNX1*$^+$ pancreatic progenitors. Discrete populations along this path can be found at each stage, as outlined in Figure 3A and detailed in Figure S7. The polyhormonal-fated cells mature in parallel, and are characterized at the end stage by their expression of *ARX*, while having no *PCSK1* or *HADH* (Figure S7). Our work thus identifies both lineage bifurcations and convergence of different paths to the same fate. Our functional assay demonstrated that *NKX6.1* can be acquired after NEUROG3, and further indicated that NEUROG3 cells forming during stage 5 are born from both NKX6.1$^+$ and NKX6.1$^-$ cells. Thereby it seems that the path in which NKX6.1 is expressed downstream of NEUROG3 is not a phenomenon restricted to stage-4 endocrine differentiation, but that both pathways (NKX6.1 expression before or after NEUROG3) are active throughout differentiation. Studies in both mouse (Nelson et al., 2007; Schaffer et al., 2013; Taylor et al., 2013) and hESC-derived pancreatic cell populations (Kelly et al., 2011; Kroon et al., 2008; Rezania et al., 2013; Russ et al., 2015) have led to the assumption that mature β cells derive from endocrine progenitors born out of NKX6.1$^+$ pancreatic progenitors. This, however, has not been unequivocally tested by specific interference solely in these cells, and our findings indicate that the onset of NKX6.1 in the endocrine progenitor population also enables β-cell differentiation. What triggers this induction deserves more investigations.

Maturation of *In-Vitro*-Derived Endocrine Cells

At stage 5 we observed high *INS* expression levels and *NKX6.1* expression in most endocrine committed cells, while *GCG* transcripts were still low or absent from about half of these cells. These findings indicate that the earliest endocrine cells express insulin alone, preceding the formation of polyhormonal cells, in agreement with the detection of insulin first in human embryos (Riedel et al., 2012). However, all endocrine cells present at the end stage of differentiation transcriptionally co-expressed multiple hormones, most notably *INS* and *GCG*. These are not background levels as they are much higher than in pancreas progenitors. A tendency to mixed α- and β-cell signatures has been seen in newly born islets but not in the adult (Wang et al., 2016), which suggests that a window of plasticity may exist until early post-natal development in humans. While *INS* was expressed at high levels in most cells, the levels of *GCG* transcript were more variable, and high expression was observed only in cells that co-expressed *ARX* and had low levels of *NKX6.1* and *PDX1*. When comparing the *in-vitro*-derived *INS*$^+$ cells with human primary α and β cells, we also detected glucagon transcripts in primary β cells, though at lower level than found in the hESC-derived β-like cells. Since most cells are not polyhormonal at the protein level, this suggests post-transcriptional regulation. A possible scenario may imply different processing of proglucagon and proinsulin. The proconvertases PCSK1 and PCSK2 are both required for proper conversion of proinsulin into insulin and C-peptide, while only PCSK2 is needed in α cells to produce glucagon. In cells co-expressing PCSK1 and PCSK2, the incretin GLP-1 would be expected instead. While *PCSK2* is expressed in almost all endocrine cells produced *in vitro*, *PCSK1* expression is variable. Thus, the *ARX*$^+$ cells with high levels of *GCG* and *INS* transcript only express *PCSK2*, and would be expected to partially process proinsulin and mature glucagon. Conversely, the β-like cells might produce mature insulin (and cleaved C-peptide) along with GLP-1 or glucagon depending of their levels of *PCSK1*. As the C-peptide targeting antibodies used in this study and most others also target proinsulin and partially processed proinsulin (Asadi et al., 2015), our analysis for C-peptide protein can likely not distinguish between these forms.

Our single-cell qPCR analysis of human islets confirmed many signature genes for the α, β, and PP cells identified in recent single-cell RNA-seq publications. Single-cell qPCR appears to be more sensitive than single-cell RNA-seq, as evidenced by, e.g., the robust detection of *MAFA* in all β cells. Finally, we identified β- and α-cell-specific expression of *SIM1*, a factor not previously described in human islet cells. Our comparison between hESC-derived *INS*$^+$ cells and human primary α and β cells reveal several genes that were expressed at similar level between mature β-like cells and human β cells, including *PTCH1*, *RUNX1T1*, *TSPAN1*, *RGS16*, *DLK1*, *PDX1*, *MNX1*, and *NKX6.1*. We further found genes enriched in primary β cells, including *IAPP*, *MAFA*, *HDAC9*, *HADH*, and *GPM6A*, while *KRT19*, *FEV*, and *PRDM16* were enriched in hESC-derived β-like cells. *GPM6A* was exclusively found in primary β cells. While *GPM6A* was not expressed in *INS*$^+$ hESC-derived cells, it was expressed in pancreatic progenitors during stage 4 and in newly born endocrine progenitors. Interestingly, GPM6A, a membrane glycoprotein, may represent a new marker that is missing from *in-vitro*-generated β-like cells that could be used as readout for maturation.

While similar hPSC models of brain and mesoderm differentiation have also identified subpopulations marking discrete maturation stages and bifurcations (Bargajea et al., 2017; Yao et al., 2016), the pancreas model shows that multiple paths can be taken *in vitro* to the same fate. It will be

important to investigate whether this is pertinent to other organs and lineages and to what extent this is conserved *in vivo*. While we compared cells produced *in vitro* with the adult cell types, comparisons with samples from human fetuses will substantiate whether the lineage specification mechanisms established *in vitro* recapitulate states that can be observed *in vivo*.

EXPERIMENTAL PROCEDURES

Human Islets and hPSCs

Human islets were obtained from Prodo Laboratories (Irvine, CA) with informed consent. hPSCs were maintained on hESC-qualified Matrigel in mTeSR-1 (see Supplemental Experimental Procedures). Experiments with human stem cells were conducted with the approval of the Science Ethical Committee of the Copenhagen region.

Differentiation of hPSCs

Differentiation of hPSCs to β-like cells was performed essentially as described in Rezania et al. (2014) or for differentiation toward polyhormonal cells, as described by Rezania et al. (2012). For details on medium composition, number of days, and factors used throughout the differentiation protocols, see Supplemental Experimental Procedures.

Sorting of Human Islets and NEUROG3-EGFP Cells

Differentiated hESCs were dissociated to single cells with TrypLE Select and human islets with Accutase (PAA Laboratories), and stained with DAPI to discriminate dead cells. Cells were sorted on a BD FACSAria Fusion cell sorter (BD Biosciences) using a 100-μm nozzle. For single-cell qPCR cells were sorted directly into 96-well PCR plates containing NP40 Detergent Surfact-Amps solution (Fisher Scientific), SUPERase-In RNase Inhibitor (Ambion), Superscript VILOcDNA Synthesis reaction mix (Invitrogen), and nuclease-free water. Plates were sealed, briefly centrifuged, and stored at −80°C until performing single-cell qPCR. Sorting of live NEUROG3-EGFP cells for subsequent culture is described in Supplemental Experimental Procedures.

Flow Cytometry and Immunofluorescence Imaging

For flow cytometry, differentiated hESCs were dissociated with TrypLE Select to a single-cell solution and fixed and stained for various intracellular markers. For immunofluorescence imaging, differentiated hESCs were fixed and stained directly in the wells of the culture plates. Supplemental Experimental Procedures provides additional details and antibody information.

Single-Cell qPCR

Reverse transcription, specific target amplification, and qPCR were performed according to the manufacturer's instructions (Fluidigm, protocol no. 68000088_G1, appendix B). Details on materials, their concentration, primer validation, initial processing of the data, quality control, and data analysis are described in Supplemental Experimental Procedures.

Statistics

For single-cell analysis, each cell was treated as a biological replicate. Error bars in the figures indicate SD or SEM as indicated in the legends.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and one data file and can be found with this article online at http://dx.doi.org/10.1016/j.stemcr.2017.08.009.

AUTHOR CONTRIBUTIONS

M.B.K.P. designed the study, performed most experiments and some bioinformatics analyses, and wrote the manuscript. A.A. carried out most bioinformatics analyses. C.I. performed immunohistochemical analysis. K.H. made the NEUROG3-GFP reporter line. C.H. and M.B.K.P. performed sorting and co-culture experiments. A.G.-B. contributed expertise in single-cell transcriptome experiments. C.H., A.G.-B., and M.H. designed the study and contributed to manuscript writing. All authors contributed to manuscript revisions.

ACKNOWLEDGMENTS

M.B.K.P. was supported by an industrial PhD grant 1355-00115B from the Innovation Fund Denmark. A.A. was supported by the Danish National Research Foundation grant DNRF 116 to the StemPhys project. Funding from the European 7th framework program project 602889 Humen and the Novo Nordisk Foundation to A.G.B. contributed to this publication. We thank Josh Brickman and Henrik Semb for discussions and comments on the manuscript. M.B.K.P., C.H., C.I., and M.H. are employees of Novo Nordisk A/S and may hold shares in the company.

Received: May 22, 2017
Revised: August 16, 2017
Accepted: August 17, 2017
Published: September 14, 2017

REFERENCES

Asadi, A., Bruin, J.E., and Kieffer, T.J. (2015). Characterization of antibodies to products of proinsulin processing using immunofluorescence staining of pancreas in multiple species. J. Histochem. Cytochem. *63*, 646–662.

Bargajea, R., Trachanaa, K., Sheltona, M.N., McGinnisa, C.S., Zhoua, J.X., Chadicka, C., Cookb, S., Cavanaugh, C., Huang, S., and Hood, L. (2017). Cell population structure prior to bifurcation predicts efficiency of directed differentiation in human induced pluripotent cells. Proc. Natl. Acad. Sci. USA *114*, 2271–2276.

Baron, M., Veres, A., Wolock, S.L., Faust, A.L., Gaujoux, R., Vetere, A., Ryu, J.H., Wagner, B.K., Shen-Orr, S.S., Klein, A.M., et al. (2016). A single-cell transcriptomic map of the human and mouse pancreas reveals inter- and intra-cell population structure. Cell Syst. *3*, 346–360.

Benitez, C.M., Qu, K., Sugiyama, T., Pauerstein, P.T., Liu, Y., Tsai, J., Gu, X., Ghodasara, A., Arda, H.E., Zhang, J., et al. (2014). An

integrated cell purification and genomics strategy reveals multiple regulators of pancreas development. PLoS Genet. *10*, e1004645.

Corish, P., and Tyler-Smith, C. (1999). Attenuation of green fluorescent protein half-life in mammalian cells. Protein Eng. *12*, 1035–1040.

Gu, G., Dubauskaite, J., and Melton, D.A. (2002). Direct evidence for the pancreatic lineage: NGN3$^+$ cells are islet progenitors and are distinct from duct progenitors. Development *129*, 2447–2457.

Heins, N., Englund, M.C.O., Sjöblom, C., Dahl, U., Tonning, A., Bergh, C., Lindahl, A., Hanson, C., and Semb, H. (2004). Derivation, characterization, and differentiation of human embryonic stem cells. Stem Cells *22*, 367–376.

Jeon, J., Correa-Medina, M., Ricordi, C., Edlund, H., and Diez, J.A. (2009). Endocrine cell clustering during human pancreas development. J. Histochem. Cytochem. *57*, 811–824.

Johnson, J.D. (2016). The quest to make fully functional human pancreatic beta cells from embryonic stem cells: climbing a mountain in the clouds. Diabetologia *59*, 2047–2057.

Kelly, O.G., Chan, M.Y., Martinson, L.A., Kadoya, K., Ostertag, T.M., Ross, K.G., Richardson, M., Carpenter, M.K., D'Amour, K.A., Kroon, E., et al. (2011). Cell-surface markers for the isolation of pancreatic cell types derived from human embryonic stem cells. Nat. Biotechnol. *29*, 750–756.

Kroon, E., Martinson, L.A., Kadoya, K., Bang, A.G., Kelly, O.G., Eliazer, S., Young, H., Richardson, M., Smart, N.G., Cunningham, J., et al. (2008). Pancreatic endoderm derived from human embryonic stem cells generates glucose-responsive insulin-secreting cells in vivo. Nat. Biotechnol. *26*, 443–452.

Lawlor, N., George, J., Bolisetty, M., Kursawe, R., Sun, L., Sivakamasundari, V., Kycia, I., Robson, P., and Stitzel, M.L. (2017). Single cell transcriptomes identify human islet cell signatures and reveal cell type-specific expression changes in type 2 diabetes. Genome Res. *27*, 208–222.

Li, J., Klughammer, J., Farlik, M., Penz, T., Spittler, A., Barbieux, C., Berishvili, E., Bock, C., and Kubicek, S. (2015). Single-cell transcriptomes reveal characteristic features of human pancreatic islet cell types. EMBO Rep. *17*, 1–10.

Löf-Öhlin, Z., Nyeng, P., Bechard, M., Hess, K., Bankaitis, E., Greiner, T., Ameri, J., Wright, C., and Semb, H. (2017). Context-specific regulation of apicobasal polarity by EGFR orchestrates epithelial morphogenesis and cellular fate. Nat. Cell Biol., In Press.

Muraro, M.J., Dharmadhikari, G., Grün, D., Groen, N., Dielen, T., Jansen, E., van Gurp, L., Engelse, M.A., Carlotti, F., de Koning, E.J.P., et al. (2016). A single-cell transcriptome atlas of the human pancreas. Cell Syst. *3*, 385–394.

Nelson, S.B., Schaffer, A.E., and Sander, M. (2007). The transcription factors Nkx6.1 and Nkx6.2 possess equivalent activities in promoting beta-cell fate specification in Pdx1+ pancreatic progenitor cells. Development *134*, 2491–2500.

Pagliuca, F.W., Millman, J.R., Gürtler, M., Segel, M., Van Dervort, A., Ryu, J.H., Peterson, Q.P., Greiner, D., and Melton, D.A. (2014). Generation of functional human pancreatic β cells in vitro. Cell *159*, 428–439.

Rezania, A., Bruin, J.E., Riedel, M.J., Mojibian, M., Asadi, A., Xu, J., Gauvin, R., Narayan, K., Karanu, F., O'Neil, J.J., et al. (2012). Maturation of human embryonic stem cell-derived pancreatic progenitors into functional islets capable of treating pre-existing diabetes in mice. Diabetes *61*, 2016–2029.

Rezania, A., Bruin, J.E., Xu, J., Narayan, K., Fox, J.K., O'Neil, J.J., and Kieffer, T.J. (2013). Enrichment of human embryonic stem cell-derived NKX6.1-expressing pancreatic progenitor cells accelerates the maturation of insulin-secreting cells in vivo. Stem Cells *31*, 2432–2442.

Rezania, A., Bruin, J.E., Arora, P., Rubin, A., Batushansky, I., Asadi, A., O'Dwyer, S., Quiskamp, N., Mojibian, M., Albrecht, T., et al. (2014). Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells. Nat. Biotechnol. *32*, 1121–1133.

Riedel, M.J., Asadi, A., Wang, R., Ao, Z., Warnock, G.L., and Kieffer, T.J. (2012). Immunohistochemical characterisation of cells co-producing insulin and glucagon in the developing human pancreas. Diabetologia *55*, 372–381.

Roark, R., Itzhaki, L., and Philpott, A. (2012). Complex regulation controls Neurogenin3 proteolysis. Biol. Open *1*, 1264–1272.

Russ, H.A., Parent, A.V., Ringler, J.J., Hennings, T.G., Nair, G.G., Shveygert, M., Guo, T., Puri, S., Haataja, L., Cirulli, V., et al. (2015). Controlled induction of human pancreatic progenitors produces functional beta-like cells in vitro. EMBO J. *34*, e201591058.

Schaffer, A.E., Taylor, B.L., Benthuysen, J.R., Liu, J., Thorel, F., Yuan, W., Jiao, Y., Kaestner, K.H., Herrera, P.L., Magnuson, M.A., et al. (2013). Nkx6.1 controls a gene regulatory network required for establishing and maintaining pancreatic Beta cell identity. PLoS Genet. *9*, e1003274.

Segerstolpe, Å., Palasantza, A., Eliasson, P., Andersson, E.-M., Andréasson, A.-C., Sun, X., Picelli, S., Sabirsh, A., Clausen, M., Bjursell, M.K., et al. (2016). Single-cell transcriptome profiling of human pancreatic islets in health and type 2 diabetes. Cell Metab. *24*, 593–607.

Soyer, J., Flasse, L., Raffelsberger, W., Beucher, A., Orvain, C., Peers, B., Ravassard, P., Vermot, J., Voz, M.L., Mellitzer, G., et al. (2010). Rfx6 is an Ngn3-dependent winged helix transcription factor required for pancreatic islet cell development. Development *137*, 203–212.

Taylor, B., Liu, F.F., and Sander, M. (2013). Nkx6.1 is essential for maintaining the functional state of pancreatic beta cells. Cell Rep. *4*, 1262–1275.

Van Der Maaten, L.J.P., and Hinton, G.E. (2008). Visualizing high-dimensional data using t-SNE. J. Mach. Learn. Res. *9*, 2579–2605.

Wang, Y.J., Schug, J., Won, K.-J., Liu, C., Naji, A., Avrahami, D., Golson, M.L., and Kaestner, K.H. (2016). Single cell transcriptomics of the human endocrine pancreas. Diabetes *65*, 3028–3038.

White, P., May, C.L., Lamounier, R.N., Brestelli, J.E., and Kaestner, K.H. (2008). Defining pancreatic endocrine precursors and their descendants. Diabetes *57*, 654–668.

Xin, Y., Kim, J., Okamoto, H., Ni, M., Wei, Y., Adler, C., Murphy, A.J., Yancopoulos, G.D., Lin, C., and Gromada, J. (2016). RNA sequencing of single human islet cells reveals type 2 diabetes genes. Cell Metab. *24*, 608–615.

Yao, Z., Mich, J.K., Ku, S., Menon, V., Krostag, A.-R., Martinez, R.A., Furchtgott, L., Mulholland, H., Bort, S., Fuqua, M.A., et al. (2016). A single-cell roadmap of lineage bifurcation in human ESC models of embryonic brain development. Cell Stem Cell *20*, 1–15.